

DOCUMENT RESUME

ED 078 097

TM 002 917

AUTHOR Smock, Richard; Crooks, Terence
TITLE A Plan for Comprehensive Evaluation of College Teaching.
INSTITUTION Illinois Univ., Urbana. Office of Instructional Resources.
PUB DATE 27 Feb 73
NOTE 20p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, Louisiana, February 25-March 1, 1973)
EDRS PRICE MF-\$0.65 HC-\$3.29
DESCRIPTORS Administrative Personnel; College Faculty; *College Instruction; Data Analysis; Effective Teaching; *Evaluation Methods; Feedback; *Program Evaluation; Student Opinion; *Teaching Methods

ABSTRACT

Functions which the evaluation of instruction are analyzed and described. In order to fulfill each of these functions, a 3 X 3 evaluation matrix incorporating three distinct "levels" of evaluation activity and three sources of evaluation information is proposed. Level 1 data will be summary data for use in campus-wide comparisons. Level 2 data will be less general, and more pointed to specific teaching attributes and classroom activities common to particular teaching units. It will be used for comparative purposes within teaching units, but, more importantly, it will serve to identify problem areas in instruction and courses. Level 3 data will be very specific feedback data aimed at pinpointing reasons for problems identified by the Level 2 evaluation, and helping in correction of such problems. The three sources for this evaluation information will be students, faculty members, and administrators. All three will have inputs into each level of the evaluation matrix. (Author/CK)

ED 078097

A PLAN FOR THE COMPREHENSIVE EVALUATION
OF COLLEGE TEACHING*

By

Richard Smock and Terence Crooks

Office of Instructional Resources
University of Illinois at
Urbana-Champaign, Illinois

U.S. DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

PART I

There has been rapidly developing pressure in recent years to provide evaluative data on college teaching; pressure from teaching faculty, from students, and from administrators. The major reasons for the pressure seem to be that teaching faculty want information which will aid them in improving their instruction, students want information to guide them in course and instructor selection, and administrators want information to guide them in pay and promotion decisions.

Much of the research related to student evaluation of teaching was recently reviewed by Costin, Greenough, and Menges (1971). Considerable effort has been expended on the development of specific rating forms, and on identifying attributes that are important components of good teaching. Musella (1966) reports that more than 3,000 studies over a sixty-six year period have attempted to isolate the variables related to effective teaching. In comparison, problems related to the systematic selection, collection, and use of evaluative data related to instruction have received less attention. Eble (1970), Miller (1972), and others have made valuable suggestions; further thought and planning along these lines seems appropriate.

* Paper presented at the Annual Meeting of the American Educational Research Association, February 27, 1973, in New Orleans, Louisiana.

TM 002 917

There are several essential points to be taken into account in any system for the evaluation of instruction. One is that the system must be feasible. For instance, while colleague classroom visitation may be an effective way of gathering information on instruction, it is not an efficient one. If such evaluation was conducted every semester, or even every year, teaching faculty could soon find themselves spending an inordinate number of hours visiting other instructors' classes - hardly a desirable solution. Efficiency as well as effectiveness needs to be considered in the development of any evaluation system.

A second point stems from the fact that most faculty members have had little or no training for teaching. For this reason it would seem inevitable (and inherently human) that they should react defensively when asked to have pay and promotion partially based on measured teaching performance. Thus, any evaluation system which does not include an adjunct service which can effectively assist faculty members as they seek to improve their instruction may be ethically questioned, and is likely to incur unnecessary resistance.

A third point is that any evaluation system must have face validity to all participants. This is a most difficult requirement to meet. Teaching faculty must see the system as providing them with helpful information, in addition to rewarding good teaching. Students must feel the effort involved on their part to be worthwhile, and administrators must have confidence in the information which they receive as a result of the evaluation system. Clearly, an essential condition for face validity of the system is that it must be flexible enough to be seen as appropriate to a great variety of instructional methods and situations.

In addition to satisfying these basic restrictions, the developer of an evaluation system must consider three major variables which apply to all evaluation. These are: who does the evaluating, what type of evaluation data is collected, and who receives and examines the results.

Much of the literature on teaching and course evaluation has focused on the information which students can provide. However, while students are valid sources for information about some aspects of instruction, a complete evaluation should include information from other sources as well. Colleagues are more appropriate judges of the quality and appropriateness of course content and related matters. Departmental administrators are in the best position to judge how well a faculty member fulfills administrative responsibilities connected with instruction, to evaluate efforts at course development and course improvement, and to evaluate contributions made by faculty members serving on departmental committees concerned with instruction. Thus, inputs from all three; students, colleagues, and administrators, seem necessary for adequate instructional evaluation.

The type of evaluative data being collected can and should vary according to the intended function of the evaluation and the people doing the evaluating. It may be helpful to identify three major types, or levels, of evaluation. The first (Level I) is general, summative evaluation, which will be concise and allow broad, general comparisons to be made across departments, but will give little or no specific information to guide instructional improvement. The second (Level II) is evaluation aimed at identifying success or failure in general areas or attributes of instruction. Most of the student evaluation forms in use today consist mainly of items to serve this purpose. The third (Level III) is detailed, course-specific evaluation aimed at providing diagnostic information about instructional problems (frequently those suggested by Level II evaluation).

Evaluative data would be useful to four audiences: instructors, students, departmental administrators, and college and campus administrators. It should not be expected that all evaluative information would be available to all audiences. For example, some types of data would be available only to the faculty member involved, and information sent to college and campus administrators would be in summary form.

In Part II of this paper, we consider the various types of data to be collected, the collection processes, and the sources of the data. In Part III, we look at the uses of the data after they have been collected.

PART II

The combination of three possible sources of data with three distinct types of data results in the data collection matrix shown below.

FIGURE 1

DATA COLLECTION MATRIX

Type of Information	Source of Information		
	Students	Colleagues	Departmental Administrators
LEVEL I Concise, general summary			
LEVEL II Specific attributes of instructors and courses			
LEVEL III Detailed diagnostic feedback			

Inputs from Students

Information for Level I evaluation needs to be comparative in nature and summary in form. Student input into this type of evaluation can be obtained from a short questionnaire, probably of no more than five to ten items which would be as applicable to an art studio course as to a chemistry lecture. If questions are to have face validity when used in courses of such wide diversity, they clearly must be very general. A request such as "Grade your instructor" would seem to have such face validity, both to the instructor and to the student, since grading is a practice with which we are all familiar. Research indicates that a very short form consisting of general questions can have acceptable face validity, and can generate reliable data for Level I evaluation (see Appendix).

Summary information of the type we are discussing is most logically collected from students at the end of a course. The instrument should have instructions which clearly indicate that the information is being collected for the purpose of rating the instructor. We stress this point because preliminary evidence gathered from over 100 university classes suggests that students are sensitive to instructions which explain the use to be made of the information. For instance, if specifically asked, they can provide very critical and diagnostic information to help improve teaching, even though their overall evaluation, as gathered by an instrument developed for Level I evaluation, may be quite favorable.

Level II evaluation by students, which is more specific than Level I but less so than Level III, relies on the ability of students as consumers of the teaching service to identify instructional problem areas, and areas of success. Implied here are instructions which make this purpose clear to the student, and items aimed at specific attributes and aspects of teaching. In order that the questionnaire have the necessary face validity to students and instructors, selection of items according to the type of instruction is necessary. Alternatively, a comprehensive

questionnaire may be used, with students being instructed not to respond to items which are not applicable to their specific course. The fact that most standardized instruments have failed in the past to make provision for the great variation in instructional practices may be partly responsible for their lack of acceptance by faculty. For instance, it is likely to be inappropriate to ask the students of an instructor teaching an art studio course to rate the organization of lectures. By the same token, there is no validity to questions which probe the quality of examination questions in an English course which has no examinations. However, both questions may have high face validity for courses using those techniques. (The surprising and disturbing fact is that we seem to be able to get data, regardless of how appropriate or inappropriate a question may be to a course.) Questions on Level II questionnaires are not normally diagnostic in the sense that they can pinpoint in detail the probable causes and remedies for instructional problems, but are rather aimed at identifying areas which may require further attention.

There is reason to keep Level I items separate from Level II items when both are included in a single questionnaire, because instructions to the students should differ since reasons for asking the questions differ. In other cases two separate questionnaire forms may be appropriate, because the summative nature of Level I evaluation makes its use at the end of the course especially apt, while Level II evaluation may well be carried out at some time during the course, thus allowing an instructor to utilize the information while teaching is in progress.

Instructors who are dissatisfied with the results of Level I evaluation, or with all or particular parts of Level II evaluation, could on their own initiative organize a Level III (detailed diagnostic) evaluation. Student contribution to this type of evaluation, intended to serve the instructor only, is not necessarily restricted to responding to forced-choice questions. In fact, open-ended questions are probably the most common method of seeking diagnostic information. For greatest benefit, open-ended questions should receive focus from information gained

in Level II evaluation. If, for instance, lectures are viewed negatively by students, specific open-ended questions concerning lectures in the course can be developed. It should be pointed out that open-ended items are not automatically Level III items, and in fact may result in information of any level, depending upon the instructions given to the students and the exact nature of the items. They are of greatest value, however, when used to provide Level III information.

There are other techniques (most of them little explored) for gathering Level III data from students. For instance, a faculty member might engage in discussions with his class, or with a few students from his class, concerning particular aspects of his teaching. Alternatively, he could arrange to have interviews with students in the course conducted by personnel from on-campus agencies charged with the task of helping instructors improve their teaching. The method chosen should be the one which is most acceptable to the instructor and most appropriate to the task. No Level III evaluation is tied to a particular point in time, and in fact such evaluations would be most effective if done early enough in a course that changes could be made and evaluated.

While up to now most of our energy has been directed towards getting Level II information from students enrolled in regular courses, any comprehensive system of instructional evaluation should include systematic Level II evaluation of instructor effectiveness related to thesis and dissertation advising.

Inputs from Colleagues

Colleagues are an important source of evaluative information about instruction. Their contribution to Level I evaluation might consist of a scaled rating of the appropriateness, comprehensiveness, and up-to-dateness of course content, textbooks, assignments, examinations, feedback, and grading. These are justifiable concerns of colleagues interested in quality instruction, and are aspects of teaching that are better judged

by fellow experts than by students. Sources of data for such judgments include examinations used in the course, assignments made to students, textbooks and reading lists, and lecture notes, as well as course descriptions and statements of objectives.

The main difference between Level II and Level I evaluation by colleagues may well be the form in which the information is presented. Thus the same sources of information as used in the Level I evaluation would form the basis of a more detailed evaluation, the results of which would be presented as a series of comments and suggestions. This more detailed Level II evaluation would be conducted concurrently with the Level I evaluation.

Colleague evaluation of a really diagnostic nature is rarely seen in college teaching. However, the development of a system in which colleague review of course content plays a part could lead to wider acceptance and use of colleagues as critics. Specific help from colleagues serving in a university unit organized to cope with instructional problems might be useful. For example, the instructor could request that a videotape be made of his teaching to be later viewed and analyzed by the instructor, perhaps with help from someone experienced at spotting areas of strength and weakness in instruction.

Inputs from Departmental Administrators

The information which could be properly allocated to the Administrator-Level I cell is summary information derived from student Level I and Level II evaluation, from colleague Level I and Level II evaluation, and from administrator Level II evaluation.

Departmental administrator input to Level II would consist of comments related to the competence of the instructor concerning the administrative aspects of teaching. Included here may be a rating of committee work directly related to teaching and teaching improvement,

time spent devising new courses and revising old ones, and time spent on departmental curricular concerns. This information is not tied to end-of-course evaluation and can be developed at a time most convenient to those concerned.

The Administrator-Level III evaluation cell is an unused cell in the matrix. Except in rare cases, it is doubtful that administrators would have truly diagnostic insights that would meet the requirements of Level III evaluation.

Part III

In the final section of this paper, we consider the uses of evaluative data after they have been collected. There are four principal users or classes of users: the faculty member whose instruction was evaluated, students, the head of department, and college and campus level administrators. We will consider each in turn.

Use of Data by the Instructor

It is, we believe, essential that all the evaluative data be made available to the faculty member directly concerned. Evaluation for the sole purpose of reward or punishment can create severe tensions, and it certainly seems inefficient not to feed back evaluative comments and give the individual concerned a chance to improve in problem areas.

Level I evaluation, whether by students, faculty colleagues, or departmental administrators, serves two main purposes in this situation. It informs the instructor of the overall opinion of the evaluators in a way that permits comparisons with other instructors, and it acts as a form of reward when the evaluation results are pleasing. It will usually be of little help in suggesting specific ways to improve.

Level II evaluation by students is likely to be important in helping to improve instruction. If problem areas do show up in such an

evaluation, the instructor has several choices. If the problem is one for which the solution is reasonably obvious, no further evaluation is likely to be needed before the instructor can attempt to improve that area of his teaching. If, however, the problem area is ill defined, or the corrective measures to be taken are not easily identified, the instructor may choose to initiate a Level III evaluation. This would normally involve further questioning of his students, or perhaps a request for help from colleagues. In many cases, help may be available from instructional service agencies within the university.

Level II evaluation by faculty colleagues concentrates mainly on technical aspects of instruction, such as course content and examinations. The instructor should be able to make use of the evaluative comments when he is revising his course, and will often wish to talk in more depth with colleagues about specific details, thus initiating Level III evaluation.

Level II evaluation by departmental administrators functions in much the same manner as Level II evaluation by colleagues. The main difference is one of substance - the areas of improvement suggested by departmental administrators will probably involve committee work and administrative matters related to instruction.

Use of Data by Students

Information on instructors or courses is useful to students when they are selecting courses. Two major types of information seem to be desirable. One type consists of summary rating data which enable comparisons among instructors to be made, and the other consists of detailed descriptive data on instructors and courses.

The first need is quite easily met, since Level I evaluation by students seems ideal for this purpose. The second need is much less readily met. It might appear that current course catalogues are adequate for this purpose, but course descriptions in such catalogues are usually out of date, give little indication of the "flavor" of the course,

and give no information on the teaching style of the instructor. As yet, no satisfactory solution to this problem seems to have been developed.

Given improved descriptions about courses and course activities, students will not usually have (or need) access to Level II or Level III evaluation data, nor to Level I evaluations by faculty colleagues and departmental administrators.

Use of Data by the Head of Department

The department head is a most important user of evaluation data. He has considerable decision-making responsibility, and is also usually the person who has to summarize the available data for submission to higher levels of administration. In some cases he will have a departmental committee to assist him in these tasks.

It seems appropriate that he should have access to all Level I and Level II data, but he will not normally receive any Level III data. In making decisions, he should be expected to use all this data, and to take into account evidence of improvement in instructional problem areas. The information to be sent on to higher levels of administration will consist of all three Level I evaluations.

Use of Data by College and Campus Administrators

The types of data which will be used by these administrators have been clearly identified in the previous section. The summary form of the data is of utmost importance, since data from many hundreds of instructors may have to be reviewed each year. It is to be expected that final recommendations by department heads will be given most weight in decision-making, but Level I student and colleague evaluations will frequently be reconsidered in all but the most obvious cases. Under some circumstances, the head of department may be asked to forward all Level II data for further detailed consideration.

CONCLUSION

We would anticipate that Level I student evaluation could be conducted each semester for all courses. Given the brevity of the Level I instrument, this seems to be feasible and would allow for feedback information to students each semester. Level II evaluation could be conducted on a time schedule thought most appropriate by departments, colleges, or individual teachers. Level III evaluation would probably seldom be initiated unless the instructor felt that there was a need for it. Given this circumstance, the number of participants in Level III evaluation might be manageable for a service unit aimed at helping an instructor improve his course and teaching.

Thus, a tri-level system of evaluation of instruction which includes input from students, colleagues, and administrators seems feasible in terms of costs (both monetary and human), and should be profitable in maximizing the education impact of instruction.

REFERENCES

F. Costin, W. T. Greenough, and R. J. Menges, "Student Ratings of College Teaching: Reliability, Validity, and Usefulness," Review of Educational Research, 41 (1971), 511-535.

Kenneth E. Eble, The Recognition and Evaluation of Teaching. Project to Improve College Teaching, (Salt Lake City), 1970.

Richard I. Miller, Evaluating Faculty Performance. San Francisco: Jossey-Bass, 1972.

Donald Musella, Teaching Effectiveness Research -- Another Approach, Council for Administrative Leadership, Albany, New York, March, 1966.

APPENDIX

At the end of this appendix is a sample copy of an evaluation form designed in accordance with many of the considerations mentioned in this paper. The first six items are Level I items, with special instructions being employed to stress their purpose, and to endeavor to persuade students to distinguish in their ratings between course content and instructor. The next ten questions are Level II items, and are focused on selected instructor attributes. They are followed by a Level III type open-ended item about the instructor. This positioning of the open-ended item, together with the instructions used, seemed to lead to very helpful comments. A similar pattern of Level II items followed by a corresponding open-ended item was used for sections on assignments and exams.

The form was used at the end of the Fall Semester, 1972, in 121 instructional sections at the University of Illinois, Urbana-Champaign. Of these 121 sections, five sections used the form but the instructor did not make the data available to us, and three sections were unable to complete all items because some items were deemed inappropriate. With these sections deleted from the sample, 113 sections remained, all but four of which were in the Physics Department. In 34 sections the instructional method was predominantly lecturing, with the instructor largely determining the content, while the remaining 79 sections were quiz-discussion sections taught by teaching assistants who had essentially no control over the course content.

The responses on each answer sheet were numerically coded, and the coded responses were key-punched into computer cards, one card per answer sheet. The cards were then fed into a computer and analyzed section by section. Whenever an item appeared to have choices ranging across a continuum from "good" to "bad," a section mean on that item was calculated. This mean could theoretically range from a high value of 5.00 down to a low value of $(6.00-J)$, where J is the number of different responses to the item. Means were calculated in this manner for seventeen items, and an

additional mean was formed by summing the means for items 3, 4, and 5, the Level I instructor items. Thus eighteen means per section were available for entry into a correlational analysis.

Some results of this analysis are shown in Table I. Let us look at some of the possible implications of these results.

Reliability

1. The extremely high intercorrelations of the Level I instructor items indicated that two or three items should be adequate for reliable Level I instructor evaluation.
2. The intercorrelations of Level I content items, while lower than those for Level I instructor items, are still high, particularly for the lecture group. The lower correlation in the case of Quiz-Discussion sections possibly resulted from smaller between-section item standard deviations, and from the fact that only five distinct courses were involved.

Validity

1. Some evidence for face validity of the form results from the very extensive usage of the form. Use was entirely voluntary for physics faculty and teaching assistants, yet well over 80% of all sections used the form.
2. Further evidence for face validity comes from the fact that only very small proportions of students failed to respond to individual questions.

3. Correlations between Level I content items and Level I instructor items are much lower than intercorrelations among the content items and among the instructor items. This indicates that the instructor and the content are being distinguished. It is pleasing to note that correlations between content and instructor items are higher for lecture sections than for discussion sections, since lecture section instructors have much more responsibility for content than do discussion section instructors.
4. It is noteworthy that the Level II items which correlate highest with Level I instructor evaluation are items 7 and 8. These specific items relate primarily to basic instructional skills rather than to instructor attitudes or enthusiasm, so the high correlations suggest that basic instructional skill is the most important factor in Level I instructor evaluation.

TABLE 1

Type of Correlation	Items	Whole Sample n = 113	Lectures n = 34	Quiz- Discussion n = 79
Intercorrelations of Level I content items	1, 2	0.852	0.894	0.748
Intercorrelations of Level I instructor items	3, 4 3, 5 4, 5	0.943 0.960 0.970	0.946 0.969 0.975	0.943 0.959 0.969
Correlations of Level I content items with sum of Level I instructor items	1, Sum 2, Sum	0.255 0.280	0.614 0.640	0.175 0.208
Correlations of Level I content items and Level I instructor sum with Level I overall item	1, 6 2, 6 Sum, 6	0.836 0.902 0.412	0.825 0.879 0.867	0.733 0.868 0.305
Correlations of Level II items with Level I instructor sum	7, Sum 8, Sum 9, Sum 10, Sum 11, Sum 13, Sum 14, Sum 16, Sum 19, Sum 22, Sum 23, Sum	0.913 0.865 0.686 0.662 0.581 0.859 0.720 0.642 0.099 0.412 0.263	0.911 0.844 0.790 0.576 0.459 0.861 0.680 0.503 0.477 0.544 0.458	0.917 0.883 0.662 0.700 0.663 0.869 0.734 0.707 -0.020 0.383 0.192

INSTRUCTIONAL EVALUATION FORM
PHYSICS - FORM C

Instructor's name _____ Course _____ Section _____

Your expected grade: A B C D E

INSTRUCTIONS:

You can help promote good instruction with your considered evaluations of certain aspects of your instructor and course. Please circle the response which most nearly expresses your opinion on each item. One student should be selected to collect all forms, place them in the envelope provided, and put the envelope in campus mail. The instructor will not see the results until he/she has submitted your final grade.

A. SUMMARY SECTION

CONTENT

Your responses to the following two items will be used in evaluating the content of the course. Please try not to let your feelings about the instructor influence your responses to these items.

1. Knowing what the course covers, if you had not already taken it (and you didn't have to take it) you would

definitely want to take it	probably want to take it	be undecided about taking it	probably not want to take it	definitely not want to take it
-------------------------------	-----------------------------	---------------------------------	---------------------------------	-----------------------------------

2. Grade the course content

A	B	C	D	E
---	---	---	---	---

INSTRUCTOR

Your responses to the following three items will be used in evaluating your instructor. When you respond, please try not to let your feelings about course content and about things which are largely out of the instructor's control (such as time of class, size of class, etc.) influence your responses.

3. The instructor's effectiveness in teaching the subject matter of this course was

excellent	very good	good	fair	poor	very poor
-----------	-----------	------	------	------	-----------

4. Assume you are considering taking another course in this department, one which your present instructor is well qualified to teach. The fact that he/she would be teaching it would

influence you very favorably toward it	influence you somewhat favorably toward it	make no difference	influence you somewhat unfavorably toward it	influence you very unfavorably toward it
---	---	--------------------	---	---

5. Grade the instructor

A	B	C	D	E
---	---	---	---	---

OVERALL

Your responses to the following item will be used in evaluating the course as a whole.

6. Overall, the course was

excellent	very good	good	fair	poor	very poor
-----------	-----------	------	------	------	-----------

B. INSTRUCTOR ATTRIBUTES

The following items are more specific items about the instructor. Please try to consider each item separately, rather than let your overall feelings about the instructor determine all the responses.

7. Presentation of course material

exceptionally clear	very clear	reasonably clear	quite unclear	very unclear
---------------------	------------	------------------	---------------	--------------

8. Ability to answer questions

consistently able to satisfy student	usually able to satisfy student	can handle most questions, but not very well	frequently unable to answer questions	item not applicable to this type of class
---	------------------------------------	--	--	--

9. Vocal delivery

speaks very clearly and distinctly	reasonably distinct	usually indistinct and/or hard to hear
---------------------------------------	---------------------	---

10. Use of blackboard

almost always clear and legible	mostly clear and legible	often faulty	very poor
------------------------------------	-----------------------------	--------------	-----------

11. Seemed to have prepared daily presentations carefully

always almost always most of the time not very often

12. The instructor brings students into classroom participation

too much an appropriate amount too little item not applicable
to this type of class

13. Ability to stimulate interest in the course material

stimulates interest stimulates interest neither stimulates reduces interest destroys interest
to a high degree nor reduces interest

14. Attitude toward students

almost always usually helpful often indifferent very negative attitude
helpful and patient and patient and/or impatient

15. Places the responsibility for learning on the student

much too heavily a little too heavily about the right amount not quite enough spoon feeds student

16. Enthusiasm in teaching the course

highly enthusiastic moderately enthusiastic shows little enthusiasm seems very unenthusiastic

C. INSTRUCTOR (comments): Write here any comments which you feel might help your instructor improve his/her teaching. You may wish to elaborate here on your responses to the previous items.

D. HOMEWORK ASSIGNMENTS AND PROBLEM SETS

These are:

17. too long about right length too short
18. too difficult of reasonable difficulty too easy
19. very helpful reasonably helpful slightly helpful useless

E. HOMEWORK ASSIGNMENTS AND PROBLEM SETS (comments): Write here any comments on the assignments which you feel might be helpful to the instructor.

F. EXAMS

These are:

20. too long about right length too short
21. too difficult of reasonable difficulty too easy
22. very fair in content okay in content unfair in content
23. very fair in grading okay in grading unfair in grading

G. EXAMS AND OTHER THINGS (comments): Comment here on exams and on other important matters not already covered (e.g., textbooks, labs). Your instructor may suggest special questions to be answered here.